Development of speech rhythm in first language: The role of syllable intensity variability

He, Lei

Abstract: The opening-closing alternations of the mouth were viewed as the articulatory basis of speech rhythm. Such articulatory cycles have been observed to highly correlate with the intensity curve of the speech signal. Analysis of the intensity variability in English monolingual children and adults revealed that (1) adults showed significantly smaller intensity variability than children, and (2) intensity variability decreased from intermediate-aged children to older children. Maturation of articulatory motor control is likely to be the main reason for the reduced variability in articulatory cycles, and hence smaller intensity variability in adults and older children.

DOI: https://doi.org/10.1121/1.5042083

Originally published at:
He, Lei (2018). Development of speech rhythm in first language: The role of syllable intensity variability. Journal of the Acoustical Society of America, 143(6):EL463-EL467.
DOI: https://doi.org/10.1121/1.5042083

Development of speech rhythm in first language: The role of syllable intensity variability

Lei He

Citation: The Journal of the Acoustical Society of America **143**, EL463 (2018); doi: 10.1121/1.5042083
View online: https://doi.org/10.1121/1.5042083
View Table of Contents: http://asa.scitation.org/toc/jas/143/6
Published by the Acoustical Society of America

Development of speech rhythm in first language: The role of syllable intensity variability

Lei He[a)]
Department of Linguistics, University of Tübingen, Wilhelmstraße 19–23, DE-72074, Tübingen, Germany
lei.he@philosophie.uni-tuebingen.de

Abstract: The opening-closing alternations of the mouth were viewed as the articulatory basis of speech rhythm. Such articulatory cycles have been observed to highly correlate with the intensity curve of the speech signal. Analysis of the intensity variability in English monolingual children and adults revealed that (1) adults showed significantly smaller intensity variability than children, and (2) intensity variability decreased from intermediate-aged children to older children. Maturation of articulatory motor control is likely to be the main reason for the reduced variability in articulatory cycles, and hence smaller intensity variability in adults and older children.

[AL]
Date Received: April 12, 2018 **Date Accepted:** May 28, 2018

1. Introduction

Speech rhythm is multidimensional,[1] yet the majority of rhythm models were heavily based on the durational dimension in some way.[1,2] The present study investigated the development of speech rhythm in first language (L1) from the perspective of intensity variability between syllables. Such intensity-based rhythm measures (intensity measures hereafter) may augment our understanding of the developmental patterns of L1 rhythm beyond the durational dimension.[3–5]

Historically, studies on speech rhythm predominantly focused on placing world languages into stress-, syllable-, or mora-timed classes based on the impressionistic judgment of isochronous grouping units.[6,7] However, failed attempts to find empirical evidence of isochrony motivated researchers to develop a number of rhythm metrics (which quantify the duration variability of the vocalic or consonantal intervals) to segregate languages traditionally labeled as stress-, syllable-, or mora-timed.[8–10] These metrics surely helped us understand how suprasegmental duration features explain some perceptually salient rhythmic differences in a number of languages.[8] However, they also oversimplified the complexity of speech rhythm by taking only the duration aspect into account, neglecting the roles of other acoustic features including intensity.[1]

What is speech rhythm? From an evolutionary viewpoint, it evolved from the pre-existing cyclical jaw movements in ancestral primates.[11–13] These movements were found to be important facial gestures in non-human primate communications.[12] In the course of human evolution, jaw cycles were coupled with vocalization: mouth opening is typically associated with sonority, and mouth closing, obstruency.[11,12] Such opening-closing gestures are temporally organized into syllable-sized units corresponding to amplitude modulations; the frequency ($\sim$5 Hz) at which these units recur is the basis of speech rhythm and is crucial to the neurological processing of the speech signal.[14,15] By calculating the spectral characteristics of the amplitude modulations, the recurring frequencies underpinning rhythmicity can be revealed.[16,17] In other words, the opening-closing alternations form the syllabic "frames," and the open and closed phases are filled with vocalic and consonantal "contents," respectively.[11] A plethora of studies on speech rhythm merely focused on the duration variability of these vocalic and consonantal contents using rhythm metrics (see Nolan and Jeon[1] and He and Dellwo[2] for reviews), including studies on L1 rhythm acquisition.[3–5] They found that younger children typically manifested less durational variability for both vocalic and consonantal intervals.

Measuring speech rhythm in terms of intensity variability was motivated by the observed phenomenon that the size of mouth aperture and the signal intensity co-

[a)]Also at: Institute of Computational Linguistics, University of Zurich, Andreasstrasse 15, CH-8050, Zurich, Switzerland.

vary: a bigger mouth opening area corresponds to a higher intensity level and vice versa.[14,18] The opening-closing gestures (i.e., the articulatory basis of speech rhythm) constantly change the vocal tract shape, and hence its filter characteristics acting upon the source signal, modifying its spectral properties and the intensity levels as a consequence. Therefore, the opening-closing cycles can be approximated by the signal intensity fluctuations. In fact, Bolton[19] in the late 19th century has astutely noted the dual roles of duration and intensity in rhythm. He used "rhythmicity" to refer to the temporal variability of a sound sequence, and "rhythmicality," the loudness variability. To investigate speech rhythm characteristics more fully, not only should we measure the rhythmicity, but also the rhythmicality. Compared to the amount of duration-based studies, intensity-based rhythm research was sporadic.[2,20–23]

In order to measure the intensity variability in the speech signal, the mean intensity of each syllable was calculated; thus, the intensity generated by each opening-closing cycle was estimated. Next, the overall and sequential intensity variabilities of each utterance were, respectively, measured in terms of the standard deviation and pairwise variability index[1,20] of syllable intensities (see Sec. 2.2 for details). Such measures evaluated the variabilities in articulatory cycles of mouth movements across an utterance.

The motivation to investigate intensity variability in both children and adults came from the fact that the articulatory motor control differs in children and adults (see Smith[24] for a comprehensive review of this topic). Particularly pertinent to the development of L1 rhythm is Schötz *et al.*[25] They examined the variability of the inter-lip aperture at the vermillion border in the midsagittal plane (which captured the joint effects of both jaw and lips) among participants with an age range of 5 to 31 years, and found that the mouth aperture variability across an utterance decreased as age increased. This means that the cyclical mouth movements vary in different age groups. Hence, the resultant intensity variability, or rhythmicality should also vary between children and adults.

The aim of this study is thus to capture such rhythmical differences in both children and adults using intensity measures (Sec. 2.2). I hypothesize that the intensity variability would be smaller in adults than in children, and amongst children intensity variability would also decrease with age. I expect that both overall and sequential measures show similar patterns between age groups, because no research, insofar as I am aware of, indicated an age effect on the overall versus sequential articulatory variabilities.

2. Method

2.1 Corpus

The corpus constructed by Polyanskaya and Ordin[3,26] was used for this study. This corpus comprises three age groups of monolingual British English-speaking children (YC, IC, and OC, see Table 1 for details) and one group of monolingual British English-speaking adults (AD, see Table 1 for details). All speakers produced 33 sentences prompted by the same pictures, hence all speech materials were, one the one hand, semi-spontaneous, and on the other hand, controlled for linguistic contents. All speech materials were recorded in mono using a Samson C01U Pro condenser microphone (Samson Technologies, Hauppauge, NY), and digitized at a sampling rate of 48 kHz and at a bit-depth of 16. Acoustic shields were used to reduce echoes and possible background noise; all speakers were sitting still in front of the microphone.[31] Annotations with different phonetic details were available; particularly relevant to this study were the syllable boundaries. Syllabifications were based on Wells,[27] and the boundary placements were dependent upon the actual phonetic realizations, rather

Table 1. Demographic details of the speakers in the corpus.

GROUP	GROUP abbreviation	n	Age range[a]	Median age[a]
Younger children	YC	12 (2 females)	4;7–5;6	5;3.5
Intermediate-aged children	IC	10 (6 females)[b]	7;4–8;5	7;10
Older children	OC	9 (2 females)	10;3–11;7	11;1
Adults	AD	10 (6 females)	25–50	42.5

[a]Age in children groups is expressed in terms of years;months.

[b]There were 21 intermediate-aged children in the original corpus. Ten were randomly selected in order to keep a balanced size comparable to other groups.

than the intended realizations.[26] More information about the corpus is available in Polyanskaya and Ordin.[3,26]

2.2 Signal processing and measurements

The intensity curve of each sentence was extracted from the original waveform following these steps: (1) The DC offset of each signal was cancelled. (2) The amplitude of each sample was squared. (3) A Kaiser-Bessel window ($\beta = 20$, sidelobe ripples attenuated by $\simeq$190 dB) with a length of 0.032 s was used to convolve the squared signal repeatedly (frame shift = 0.008 s, 75% between-frame overlap). (4) For each windowed frame, the sum of squares (SoS) of the sample values was computed and substituted in $10 \times \log\{[\mathrm{SoS}/(2 \times 10^{-5})]^2/0.032\}$ to obtain the intensity level (unit: dB re 20 μPa) in each particular frame. (5) The intensity curves of all sentences were finally linearly normalized such that the new average intensity equated to 65 dB (re 20 μPa), while maintaining the shapes of the original curves.

Intensity measures were then calculated from the intensity curves. For a sentence with n syllables, the mean intensity of each syllable (I_i, $i \leq n \in \mathbb{Z}^+$) was obtained; this gives an estimate of intensity generated by each articulatory cycle corresponding to a syllable. To capture the intensity variability of this sentence, the standard deviation (stdev-I) and pairwise variability index [$\mathrm{PVI\text{-}I} = (|I_1 - I_2| + |I_2 - I_3| + \cdots + |I_{n-1} - I_n|)/(n - 1)$][1,20] were calculated. They accounted for the overall and sequential intensity variability across an utterance, respectively.

2.3 Statistical analysis

Linear mixed models (by-item design, i.e., repeated for SENTENCE) fitted by maximum likelihood were used for data analysis. The stdev-I and PVI-I were modeled as dependent variables. In a full model, GROUP (YC, IC, OC, and AD, see Table 1) was modeled as the fixed factor, and SENTENCE was modeled as the random intercept. In a reduced model, GROUP was eliminated. To test the effect of GROUP, a likelihood ratio χ^2 test was run between a full model and a reduced model; a significant χ^2-statistic would indicate that the GROUP effect was significant. *Post hoc* comparisons between groups were made using least square means. The Tukey method was used to adjust p-values.

3. Results

The GROUP effect was significant for both stdev-I ($F_{[3,1317]} = 12.77$, $p < 0.0001$) and PVI-I ($F_{[3,1317]} = 16.12$, $p < 0.0001$). The results of likelihood ratio tests for model comparisons are presented in Table 2. *Post hoc* comparisons (see Fig. 1) indicated a general developmental pattern of rhythmicality from children to adults; nevertheless, the differences between YC and IC as well as YC and OC were not significant for both stdev-I and PVI-I.

4. Discussion

The results generally conformed to the hypothesis that adults manifest smaller suprasegmental intensity variability than children, and the pattern was similar for both stdev-I and PVI-I (see Fig. 1). Moreover, among children groups, OC was significantly smaller than IC in terms of both stdev-I and PVI-I (though the differences between YC and IC/OC were not significant). Given that a strong association exists between the mouth aperture size and the intensity curve,[14,18] one can reasonably argue that the general decrease of intensity variability across an utterance from childhood to adulthood, to a large extent, is a consequence of the decremented inter-lip aperture variability en route to maturation.[25] Further support to this claim is offered by Smith and Zelaznik[28] who investigated how the functional synergies for speech motor coordination developed from children to adults. They discovered that children exhibited less

Table 2. Results of likelihood ratio tests for model comparisons between the full and GROUP-reduced models. The χ^2-statistics were significant for both stdev-I and PVI-I. The full models showed smaller AICs and BICs, suggesting better model fits.

	Df	AIC	BIC	−logLikelihood ratio	Deviance	$\chi^2_{[\mathrm{Df}]}$	p
(i) Dependent variable: stdev-I							
GROUP-reduced model	3	6659.1	6674.7	3326.5	6653.1		
Full model	6	6627.2	6658.5	3307.6	6615.2	$37.8_{[3]}$	$\ll$0.0001
(ii) Dependent variable: PVI-I							
GROUP-reduced model	3	7381.5	7397.1	3687.8	7375.5		
Full model	6	7339.9	7371.2	3664.0	7372.9	$47.6_{[3]}$	$\ll$0.0001

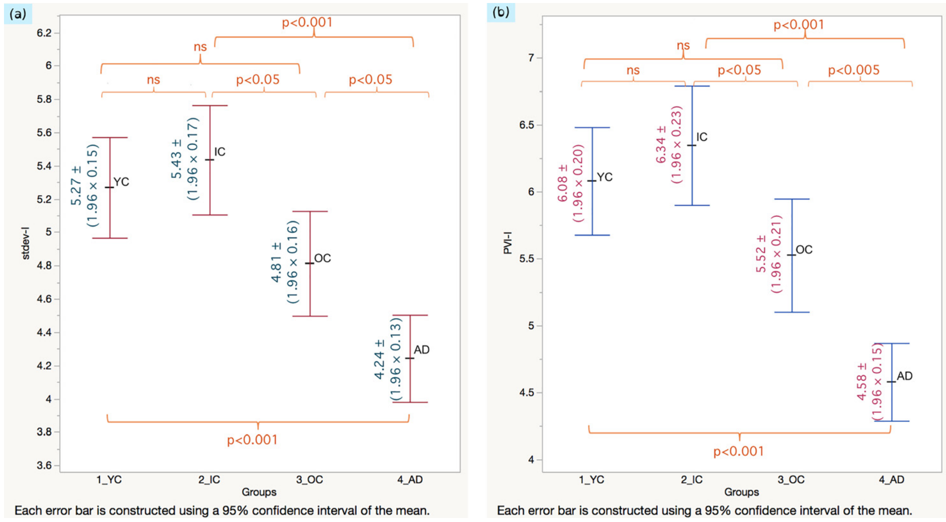

Fig. 1. (Color online) Error bar plots showing the differences between the four groups (YC, IC, OC, and AD) in terms of stdev-I (a) and PVI-I (b). The means and 95% confidences intervals (1.96 × standard errors) are shown next to the error bars. The p-values of *post hoc* comparisons are marked; "ns" means non-significant ($p > 0.05$).

consistent motion relationship among the upper lip, lower lip, and jaw in sentence production. In contrast, adults showed more regular coupling patterns among these articulators. Such developmental patterns generally conformed to the patterns of measured rhythmicality between children (IC and OC) and adults of the present study. On the one hand, immature neuro-motor control of the articulators in children may be the reason for their high articulatory and hence intensity variability; on the other hand, the developing craniofacial architecture in children may constrain the biomechanical properties of the articulators from manifesting more regular and consistent articulatory cycles typically found in adults.[28] The results implied that the development of the articulatory motor control from childhood to adulthood may result in reduced variability in the articulatory cycles underpinning speech rhythm, which may be measurable using intensity measures. The exact relationships between articulatory movements and intensity variations across different age groups are subject to further research.

Nevertheless, the intensity variability of the YC group was unexpected. Studies on the development of speech motor control[25,28] showed that the five-year olds exhibited the least regular articulatory coordination, yet their intensity variability was not significantly different from other children groups (IC and OC). This suggests that articulatory regularity may not be the only driving force for intensity variations. Another source contributing to intensity may be the aerodynamics in speech production. Stathopoulos and Weismer[29] found that the intraoral air pressure was not significantly different between 4 to 8 and 10 to 12 years old English speaking children. This might explain the non-significant differences between YC and IC/OC in the intensity measures, since the overall aerodynamics were similar across these ages. How aerodynamic characteristics and articulatory movements actually interact to influence intensity dynamics as a function of age is subject to more in-depth research.

The results of this study complement our understanding of rhythm acquisition in L1. Comparing the results from Polyanskaya and Ordin[3] using duration-based rhythm measures, we can observe an opposite pattern between children and adults: duration variabilities of consonantal and vocalic intervals increase as a function of age. This suggests that the timing of the vocalic/consonantal contents and the intensity organization of syllable frames are two independent processes, even though both of them are the acoustic outcomes of the same speech motor commands. Whether the perceived differences in rhythm between children and adults are more due to the duration dimension or the intensity dimension or both is subject to more research.

For further research, it is important to study speech rhythm development using languages with different phonological complexities (which requires different degrees of articulatory control), and test whether similar results would replicate. Moreover, to better understand the role of articulatory movements in the production of speech rhythm in different populations (e.g., speakers of different age groups, different languages, or having different forms of speech pathologies), it is imperative to record articulatory trajectories to characterize speech rhythm, possibly by analyzing the coherence spectrum[14,30] between the signals from the articulatory and the acoustic domains.

Acknowledgments

This study was benefited from an Early Postdoc Mobility Grant of the Swiss National Science Foundation (Grant No. P2ZHP1_178109). Many thanks go to Mikhail Ordin and Leona Polyanskaya for making their corpus available, and helpful explanations of data collection details. I acknowledge support by Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of University of Tübingen.

References and links

[1] F. Nolan and H.-S. Jeon, "Speech rhythm: A metaphor?," Philos. Trans. R. Soc. B **369**, 20130396 (2014).
[2] L. He and V. Dellwo, "The role of syllable intensity in between-speaker rhythmic variability," Int. J. Speech Lang. Law **23**, 243–273 (2016).
[3] L. Polyanskaya and M. Ordin, "Acquisition of speech rhythm in first language," J. Acoust. Soc. Am. **138**, EL199–EL204 (2015).
[4] M. Ordin and L. Polyanskaya, "Development of timing patterns in first and second languages," System **42**, 244–257 (2014).
[5] E. Payne, B. Post, L. Astruc, P. Prieto, and M. del M. Vanrell, "Measuring child rhythm," Lang. Speech **55**, 203–229 (2012).
[6] A. Lloyd James, *Speech Signals in Telephony* (Sir I. Pitman, London, UK, 1940).
[7] D. Abercrombie, *Elements of General Phonetics* (Edinburgh University Press, Edinburgh, UK, 1967).
[8] F. Ramus, M. Nespor, and J. Mehler, "Correlates of linguistic rhythm in the speech signal," Cognition **73**, 265–292 (1999).
[9] E. L. Low, E. Grabe, and F. Nolan, "Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English," Lang. Speech **43**, 377–401 (2000).
[10] E. Grabe and E. L. Low, "Durational variability in speech and rhythm class hypothesis," in *Laboratory Phonology VII*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin, Germany, 2002), pp. 514–546.
[11] P. F. MacNeilage, "The frame/content theory of evolution of speech production," Behav. Brain Sci. **21**, 499–546 (1998).
[12] A. A. Ghazanfar, C. Chandrasekaran, and R. J. Morrill, "Dynamic, rhythmic facial expressions and the superior temporal sulcus of macaque monkeys: Implications for the evolution of audiovisual speech," Eur. J. Neurosci. **31**, 1807–1817 (2010).
[13] R. J. Morrill, A. Paukner, P. F. Ferrari, and A. A. Ghazanfar, "Monkey lipsmacking develops like the human speech rhythm," Dev. Sci. **15**, 557–568 (2012).
[14] C. Chandrasekaran, A. Trubanova, S. Stillittano, A. Caplier, and A. A. Ghazanfar, "The natural statistics of audiovisual speech," PLoS Comput. Biol. **5**, e1000436 (2009).
[15] C. Chandrasekaran and A. A. Ghazanfar, "Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus," J. Neurophys. **101**, 773–788 (2009).
[16] S. Tilsen and K. Johnson, "Low-frequency Fourier analysis of speech rhythm," J. Acoust. Soc. Am. **124**, EL34–EL39 (2008).
[17] S. Tilsen and A. Arvaniti, "Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages," J. Acoust. Soc. Am. **134**, 628–639 (2013).
[18] L., He and V. Dellwo, "Between-speaker variability in temporal organizations of intensity contours," J. Acoust. Soc. Am. **141**, EL488–EL494 (2017).
[19] T. L. Bolton, "Rhythm," Am. J. Psychol. **6**, 145–238 (1894).
[20] E. L. Low, "Prosodic prominence in Singapore English," Doctoral dissertation, University of Cambridge, Cambridge, UK (1998).
[21] L. He, "Syllabic intensity variations as quantification of speech rhythm: Evidence from both L1 and L2," in *Proceedings of Speech Prosody*, Shanghai, China (2012), pp. 466–469.
[22] R. Fuchs, *Speech Rhythm in Varieties of English* (Springer, Singapore, 2016).
[23] E. Ferragne and F. Pellegrino, "Le rythme dans les dialectes de l'anglais: Une affaire d'intensité?" ("The rhythm in the dialects of English: A matter of intensity?"), in *Actes de Journées d'Étude de la Parole* (Avignon, France, 2008), pp. 1678/1–4.
[24] A. Smith, "Development of neural control of orofacial movements for speech," in *The Handbook of Phonetic Sciences*, 2nd ed., edited by W. J. Hardcastle, J. Laver, and F. E. Gibbon (Wiley-Blackwell, Oxford, UK, 2010), pp. 251–296.
[25] S. Schötz, J. Frid, and A. Löfqvist, "Development of speech motor control: Lip movement variability," J. Acoust. Soc. Am. **133**, 4210–4217 (2013).
[26] M. Ordin and L. Polyanskaya, "A database for L1 rhythm research," The IRIS Repository, https://www.iris-database.org/iris/app/home/detail?id=york%3a853420 (Last viewed May 17, 2018).
[27] J. C. Wells, *Longman Pronunciation Dictionary*, 3rd ed. (Pearson Education Limited, Essex, UK, 2008).
[28] A. Smith and H. N. Zelaznik, "Development of functional synergies for speech motor coordination in childhood and adolescence," Dev. Psychobiol. **45**, 22–33 (2004).
[29] E. L. Stathopoulos and G. Weismer, "Oral airflow and pressure during speech production: A comparative study of children, youths and adults," Folia Phoniatr. Logop. **37**, 152–159 (1985).
[30] A. M. Alexandrou, T. Saarinen, J. Kujala, and R. Salmelin, "A multimodal spectral approach to characterize rhythm in natural speech," J. Acoust. Soc. Am. **139**, 215–226 (2016).
[31] M. Ordin (private communication).